

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Seiring berkembangnya kecanggihan teknologi internet dalam pengembangan *mobile application* dan keamanan dalam pembayaran melalui elektronik, kegiatan jual beli melalui *e-commerce* menjadi aktivitas keseharian. Indonesia menjadi pasar yang sangat menjanjikan dengan melihat fakta yang diambil dari *ecommerceiq.asia* pada bulan Maret 2017 yang lalu, jumlah *web traffic* dari salah satu pelaku *e-commerce* di Indonesia, yaitu Tokopedia mencapai hingga 51 juta kunjungan, selengkapnya dapat dilihat di Lampiran 1. Melihat fakta yang terjadi kebutuhan dari *e-customer* tentu saja semakin meningkat dan menginginkan kualitas pelayanan yang baik kapanpun dibutuhkan. Di sisi lain, *online retailer* menginginkan untuk bisa memahami *customer* dan mampu memprediksi apa yang *customer* butuhkan.

Di waktu yang sama, dengan melihat besarnya potensi penerapan dalam ilmu pengetahuan yang dapat diaplikasikan, *Web Usage Mining* (WUM) menjadi salah satu area pengaplikasian yang aktif untuk diteliti, menawarkan berbagai teknik *data mining* untuk mengekstrak informasi berharga melalui kebiasaan *user* yang tersimpan di dalam *server log* [1]. Aktivitas yang dilakukan oleh *user* ketika mengunjungi sebuah *website* terekam dengan baik di dalam *server log*, data ini yang menjadi data sumber utama WUM dalam mencari informasi berharga.

Mengidentifikasi *session* pengguna *e-commerce* yang berakhir dengan pembelian, menjadi topik yang hangat dibicarakan saat ini [2]. Sebuah *session* yang teridentifikasi *browsing session* dapat memicu sistem rekomendasi yang mampu meningkatkan kemungkinan untuk membeli. Selain itu, sebuah *session* yang diprediksi sebagai *buying session* memungkinkan untuk sistem rekomendasi menambahkan rekomendasi produk untuk diprediksi [2]. Salah satu hal fundamental yang dibutuhkan sistem rekomendasi adalah mengetahui bagaimana *user profile* dari *user* yang mengunjungi sebuah *website* tersebut.

Pada pengerjaan Tugas Akhir ini, penulis mengidentifikasi terdapat dua permasalahan yang akan diselesaikan, yakni memprediksi *session* dalam *e-commerce website* dan juga mengelompokkan *e-customer* berdasarkan kesamaan pola navigasi. Penulis mengusulkan dengan menggunakan dari teknik *data mining* yang cukup populer yaitu *Support Vector Machine* (SVM) *Classification* dan *Enhanced Fuzzy C-Means Clustering*. SVM merupakan teknik yang *powerful* dalam membentuk sebuah model untuk prediksi berdasarkan *supervised classification*, yang memungkinkan untuk mengenali pola yang memiliki kompleksitas tinggi dengan dimensi data yang besar. SVM juga telah dibuktikan untuk membantu *customer* dalam *online shopping* melalui teknik *recommendation system*, dan juga mendesain *intelligent software agent* yang mampu menyeleksi dan menghubungkan produk yang sesuai dengan *customer* [1]. Sedangkan *Fuzzy Clustering* telah diteliti sebagai teknik yang sesuai dalam menangani ambiguitas data, karena memungkinkan pembentukan *overlapping cluster* dan juga pemberian derajat keanggotaan data [3].

1.2 Perumusan Masalah

Adapun rumusan masalah yang akan dilaksanakan dalam pengerjaan Tugas Akhir ini yang didapat dari latar belakang masalah diatas adalah sebagai berikut:

1. Bagaimana memprediksi antara *Buying Session* dan *Browsing Session* dari setiap *session* pengguna di *server log* dengan menggunakan metode *Support Vector Machine*?
2. Bagaimana mendapatkan *user profile* dari kelompok *buying session* dan *browsing session* yang dikelompokkan berdasarkan pola navigasi dari *session* pengguna yang serupa dengan menggunakan *Enhanced Fuzzy C-Means*?
3. Bagaimana hasil dari *Enhanced Fuzzy C-Means* yang diterapkan pada pembentukan *user profiles*?

Untuk memperkecil cakupan dari Tugas Akhir ini diperlukan batasan – batasan, adapun batasan masalah yang diambil untuk Tugas Akhir ini adalah sebagai berikut:

1. Data yang digunakan adalah data dari *Recsys challenge* yang disediakan oleh yochoose (<http://2015.recsyschallenge.com/>) sejumlah 1.500.000 baris.
2. Penelitian yang dilakukan saat ini masih bersifat *offline*. Yaitu dikembangkannya sistem yang bersifat offline, dan menggunakan data yang bersifat tetap.
3. Dalam pembentukan model klasifikasi yang menggunakan SVM, *kernel polynomial* tidak dilakukan karena komputasi yang sangat berat.

1.3 Tujuan

Dari rumusan masalah yang ada, didapatkan tujuan dari penyelesaian Tugas Akhir ini, yaitu:

1. Menerapkan model yang paling tepat dalam membedakan antara *buying session* dan *browsing session* dari setiap *session* pengguna pada *server log*.
2. Mengetahui *user profile* dari pengguna yang dikelompokkan berdasarkan pola navigasi dari *session* pengguna yang serupa.
3. Mengetahui analisis dan hasil yang didapatkan dari penerapan *Enhanced Fuzzy C-Means* pada pembentukan *user profiles*.

1.4 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam menyelesaikan Tugas Akhir ini adalah sebagai berikut:

1. Studi Literatur

Diawali dengan membaca referensi yang berkaitan dengan metode yang digunakan dalam Tugas Akhir ini. Referensi acuan utama yang digunakan dalam Tugas Akhir ini berjudul “*Web Log Data Analysis by Enhanced Fuzzy C Means Clustering*” yang ditulis oleh V. Chitraa dan Antony Selvados Thanamani pada tahun 2014. Pada *paper* diatas membahas tentang analisis implementasi dari *web*

log dalam *website* yang menggunakan algoritma *Enhanced Fuzzy C means* sebagai metode *clustering* dan *Support Vector Machine* sebagai klasifikasi. Selain membaca referensi yang berkaitan dengan metode utama yang dipakai, metode lain yang mendukung metode utama juga berawal dari membaca dari berbagai referensi.

2. Pengumpulan Data

Data untuk Tugas Akhir ini didapatkan melalui tantangan yang disediakan oleh Yoochoose di *RecSys Challenge 2015* sebagai sumber utama untuk pengerjaan *Web Usage Mining*.

3. Konsultasi

Melakukan bimbingan kepada dosen pembimbing untuk mengetahui langkah selanjutnya yang dapat dilakukan agar pengerjaan Tugas Akhir bisa diselesaikan sesuai waktu yang ditentukan. Bimbingan juga dilakukan untuk memahami konsep yang kurang dimengerti dari metode yang digunakan dalam Tugas Akhir sehingga pengerjaan tetap berjalan lancar.

4. Perancangan dan Implementasi Sistem

Setelah melakukan identifikasi masalah dan studi literatur, maka dibuatlah perancangan sistem agar mempermudah dan memperjelas tahap implementasi sistem. Dan implementasi sistem menggunakan 120.000 *log session* data latihan yang digunakan untuk membentuk sebuah model klasifikasi.

5. Pengujian dan Analisis Hasil

Dilakukannya analisis untuk mendapatkan model dengan akurasi yang paling tinggi dan mencari jumlah kaster yang terbaik untuk bisa mengelompokkan pola akses *user* yang serupa.

6. Pengambilan Kesimpulan dan Penyusunan Buku

Mengambil kesimpulan berdasarkan hasil uji coba dan analisis yang telah dilakukan, dan penyusunan buku yang berisi laporan serta dokumentasi selama pengerjaan tugas akhir.

1.5 Sistematika Penulisan

Penulisan Tugas Akhir ini disusun dalam beberapa bab yang dari setiap bab, berisi data – data sebagai berikut:

1. BAB 1 PENDAHULUAN

Berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan yang ingin dicapai dan metodologi penyelesaian masalah dalam pengerjaan Tugas Akhir ini.

2. BAB 2 DASAR TEORI

Bab ini berisi tentang teori yang mendukung dalam penyusunan Tugas Akhir, teori yang dibahas menyangkut metode *Enhanced Fuzzy C-Means* dan *Support Vector Machine*, yaitu metode yang digunakan untuk *clustering* dan klasifikasi.

3. BAB 3 PERANCANGAN SISTEM

Bab ini berisi penjelasan tentang sistem yang akan dibuat untuk Tugas Akhir ini. Menjelaskan tentang *flowchart* dari awal pembangunan model sampai proses percobaan. Penjelasan mengenai *dataset* yang dipakai secara detail dijelaskan pada bagian awal dari bab ini.

4. BAB 4 PENGUJIAN DAN ANALISIS

Bab ini berisi hasil pengujian yang berisi akurasi dari sistem beserta perubahan data saat dilakukan skenario percobaan. Kemudian analisis hasil dari yang dicapai juga terdapat pada bab ini.

5. BAB 5 KESIMPULAN DAN SARAN

Kesimpulan dari penyusunan Tugas Akhir berada pada bab ini, kemudian untuk perbaikan yang dapat dilakukan untuk penelitian selanjutnya disertakan dalam saran bersamaan dengan kesimpulan.